

Lessons learned from predicting the structure of over 100 Klebsiella phage RBPs

Dimitri Boeckaerts, Michiel Stock, Bernard De Baets, Rafael Sanjuan, Pilar Domingo-Calap, Yves Briers

Predicting phage RBP structures using AlphaFold

Over the last year, Deepmind's AlphaFold2 system has taken the structural biology field by a storm by enabling accurate predictions of protein structures with artificial intelligence. In the phage field too, AlphaFold can have a significant impact in phage therapy applications or for phage engineering efforts.

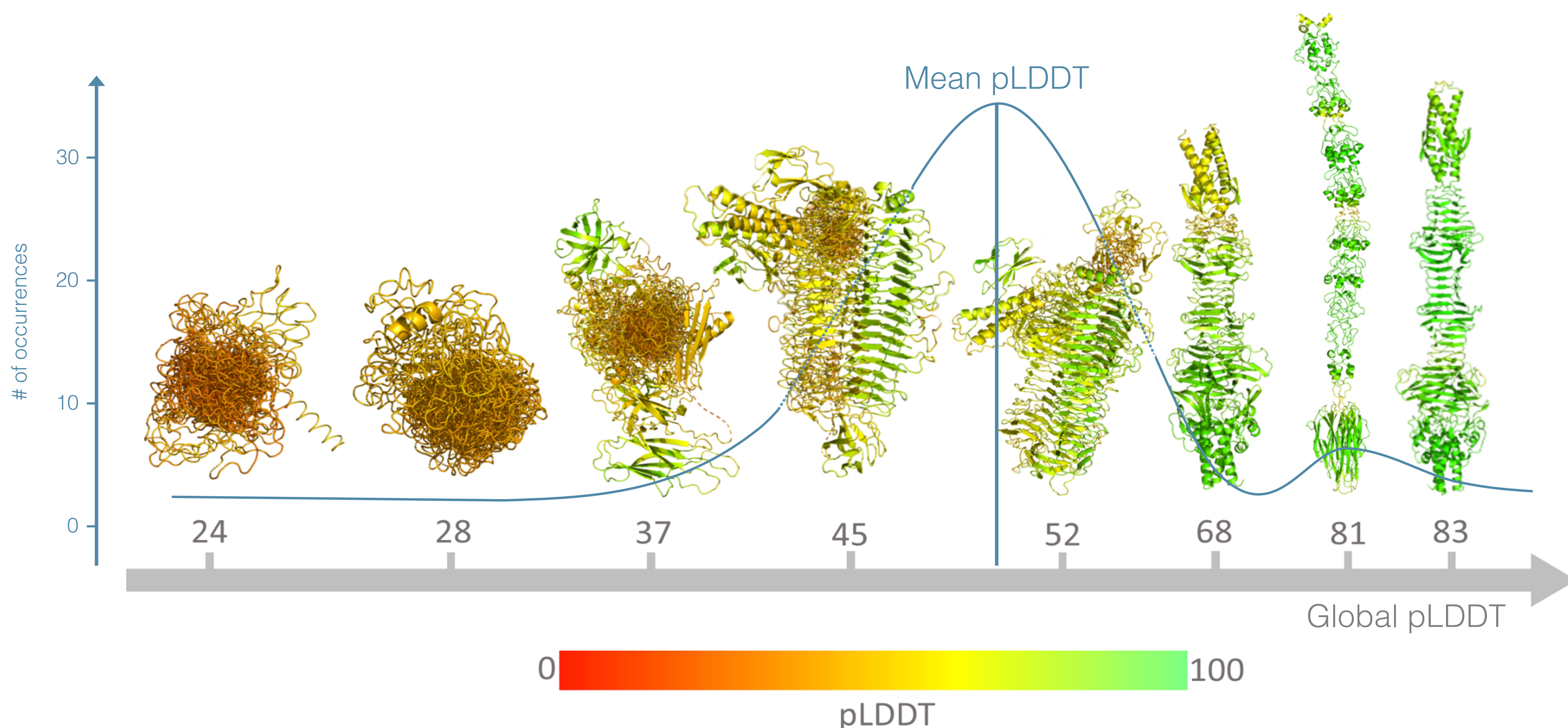
One particularly interesting class of phage proteins are receptor-binding proteins (RBPs), which can be used in phage engineering to adjust host specificity. However, RBP structures are not widely available, hampering structure-guided design and engineering of RBPs.

Research aim

We have used AlphaFold to predict the structure of over 100 RBPs of Klebsiella phages. Our experiments show that predicting the trimeric structure of phage RBPs can be challenging but overall shed a new light on proteins with missing annotation. Specifically, the accuracy of AlphaFold's predictions can be highly variable, ranging from structures that are predicted nearly perfect, to structures that completely fail to be predicted.

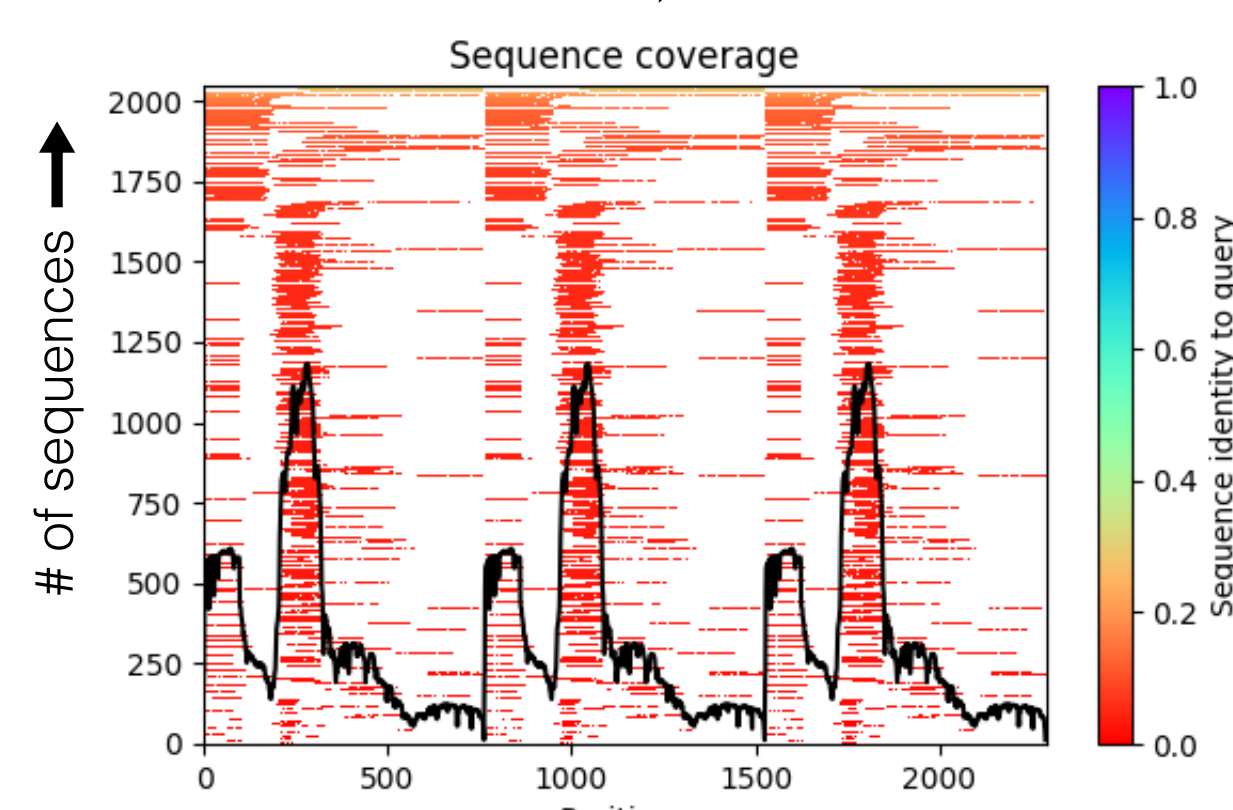
With this work, we aim to contribute to the broad and swift use of AlphaFold by phage researchers to study phage proteins in a comprehensive but attentive manner.

Predicting the structure of phage RBPs using AlphaFold yields highly variable pLDDT scores

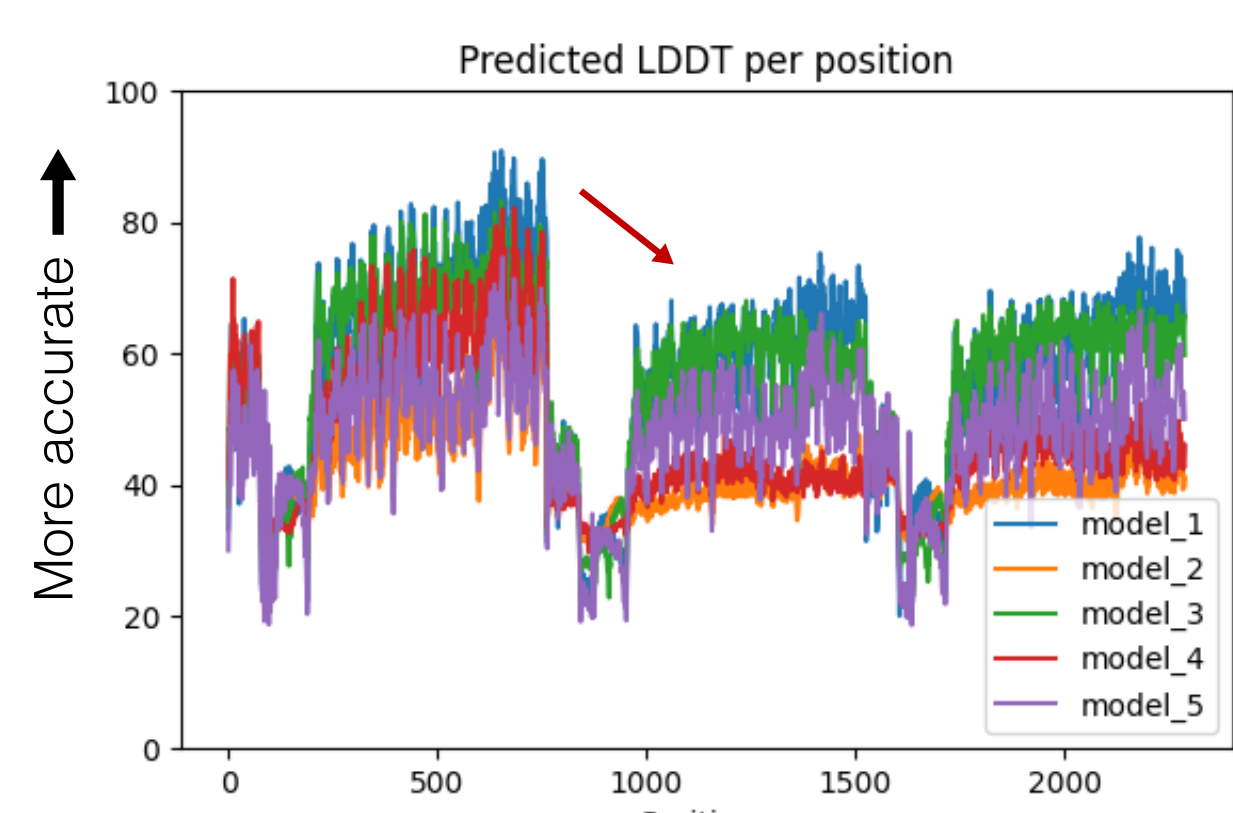
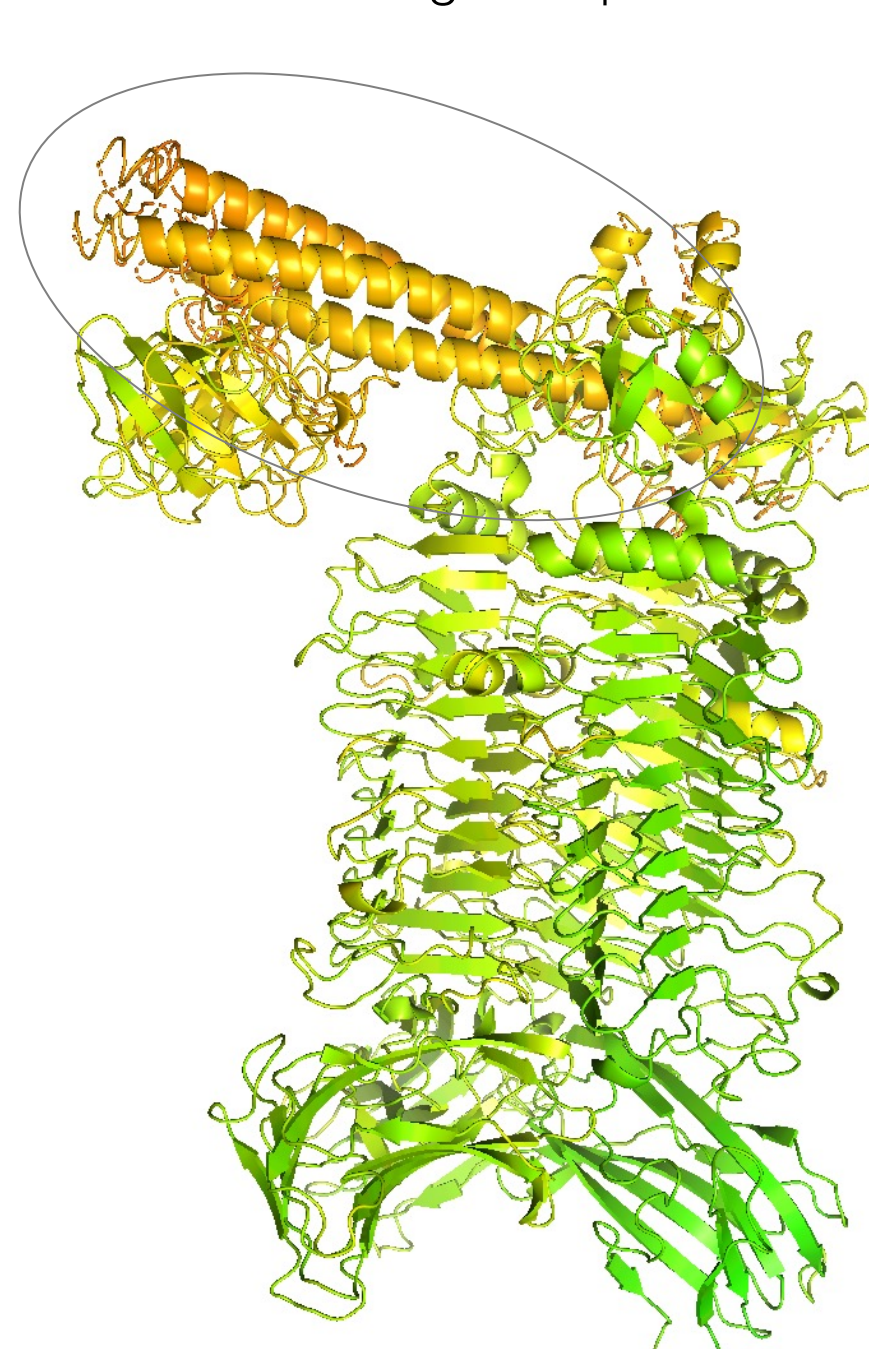


Trimeric predictions can be challenging

High sequence coverage is not sufficient for accurate structures, and vice versa.



Proper mutual orientation of the subdomains can be challenging for AlphaFold.



First monomers are typically predicted more accurately than other monomers.

Lessons learned

- Phage RBP structures are predicted with varying accuracy. More experimental data are needed to improve predictions.
- Sufficient computing power (HPC) is needed to consistently enable trimeric structure predictions.
- Typically, the most difficult to properly predict is the mutual orientation of the phage RBP subdomains.
- In trimeric structure predictions, the first monomer is often the most accurately predicted.
- Good sequence coverage to the query doesn't always lead to an accurately predicted structure of that query.
- Increasing AlphaFold's recycles can improve predictions, albeit not consistently.